Different control conditions can produce different effect estimates in psychotherapy trials for depression

Michopoulos, Ioannis; Furukawa, Toshi A; Noma, Hisashi; Kishimoto, Sanae; Onishi, Akira; Ostinelli, Edoardo G; Ciharova, Marketa; Miguel, Clara; Karyotaki, Eirini; Cuijpers, Pim

published in
Journal of clinical epidemiology
2021

DOI (link to publisher)
10.1016/j.jclinepi.2020.12.012

document version
Publisher's PDF, also known as Version of record

Journal of Clinical Epidemiology 132 (2021) 59−70

Journal of Clinical Epidemiology

REVIEW

Different control conditions can produce different effect estimates in psychotherapy trials for depression

Ioannis Michopoulos[a,1,*], Toshi A. Furukawa[b,1], Hisashi Noma[c], Sanae Kishimoto[d], Akira Onishi[e], Edoardo G. Ostinelli[f], Marketa Ciharova[g], Clara Miguel[g], Eirini Karyotaki[h,g], Pim Cuijpers[g]

[a]*Eating Disorders Unit, 2nd Department of Psychiatry, Medical School, National and Kapodistrian University of Athens, 'Attikon' University Hospital, Athens, Greece*
[b]*Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan*
[c]*Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan*
[d]*Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine / School of Public Health, Kyoto, Japan*
[e]*Department of Rheumatology and Clinical Immunology, Kobe University Graduate School of Medicine, Hyogo, Japan*
[f]*Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK; Department of Psychiatry, University of Oxford, Oxford, UK*
[g]*Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands*
[h]*Department of Global Health and Social Medicine, Harvard Medical School, Boston MA, USA*

Accepted 8 December 2020; Published online 15 December 2020

Abstract

Objectives: Control conditions' influence on effect estimates of active psychotherapeutic interventions for depression has not been fully elucidated. We used network meta-analysis to estimate the differences between control conditions.

Study Design and Setting: We have conducted a comprehensive literature search of randomized trials of psychotherapies for adults with depression up to January 1, 2019 in four major databases (PubMed, PsycINFO, Embase, and Cochrane). The network meta-analysis included broadly conceived cognitive behavior therapies in comparison with the following control conditions: Waiting List (WL), No Treatment (NT), Pill Placebo (PillPlacebo), Psychological Placebo (PsycholPlacebo).

Results: 123 studies with 12,596 participants were included. The I-squared was 55.9% (95% CI: 45.9%; to 64.0%) (moderate heterogeneity). The design-by-treatment global test of inconsistency was not significant ($P = 0.44$). Different control conditions led to different estimates of efficacy for the same intervention. WL appears to be the weakest control (odds ratio of response against NT = 1.93 (1.30 to 2.86), PsycholPlacebo = 2.03 (1.21 to 3.39), and PillPlacebo = 2.66 (1.45 to 4.89), respectively).

Conclusions: Different control conditions produce different effect estimates in psychotherapy randomized controlled trials for depression. WL was the weakest, followed by NT, PsycholPlacebo, and PillPlacebo in this order. When conducting meta-analyses of psychotherapy trials, different control conditions should not be lumped into a single group.

Keywords: Control conditions; Depression; Network meta-analysis; Psychotherapy; RCT

Author's contribution: I.M. took part in the conceptualization and in writing the original draft and the review; T.A.F. took part in the conceptualization, methodology, supervision, and writing the original draft and the review; H.N. took part in the formal analysis; S.K., A.O., E.O., M.C., C.M., and E.K. took part in the acquisition of data and analysis and interpretation of data; P.C. took part in the conceptualization, methodology, supervision, and writing the original draft. All authors have approved the version to be submitted.

Toshiaki Furukawa reports personal fees from Mitsubishi-Tanabe, MSD and Shionogi, and a grant from Mitsubishi-Tanabe, outside the submitted work; TAF has a patent 2018-177688 pending. Ioannis Michopoulos has received consultancy fees from Lundbeck. Pim Cuijpers reports support for unrelated grants from the European Commission, and ZonMw, expense allowances for his membership of the Board of Directors of the "Mind.nl," "Fonds Psychische Gezondheid," and "Korrelatie," for being Chair of the PACO Committee of the "Raad voor Civiel-militaire Zorg en Onderzoek" of the Dutch Ministry of Defense, and for being Chair of the Mental Health Priority Area of the Welcome Trust in London, UK, and he receives royalties for books and for occasional workshops and invited addresses. Edoardo Ostinelli has received research and consultancy fees from Angelini Pharma. All other authors report no financial or other conflicts of interest.

[1] Equal contribution.

* Corresponding author. Eating Disorders Unit, 2nd Department of Psychiatry, Medical School, National and Kapodistrian University of Athens, "Attikon" University Hospital, Rimini 1, 12462, Athens, Greece. Tel: +30 2105832430; fax: +302105326453.

E-mail address: imihopou@med.uoa.gr (I. Michopoulos).

https://doi.org/10.1016/j.jclinepi.2020.12.012

1. Introduction

More than 300 million people are affected by depression globally [1]. Its prevalence has substantially increased for the last 30 years, due to population growth and aging [2]. In high-income countries, it is estimated that depression will be the main cause of disability by the year 2030 [1]. Depression is a public−health priority [3−5]. Psychotherapy is one of the two major interventions for depression [6,7]. There is accumulating evidence that psychotherapeutic interventions can contribute a lot or even be the major therapeutic approach for several mental disorders [8,9]. Several psychological interventions have been developed and have shown efficacy for the treatment of depression [10].

Such recommendations are based on randomized evidence. A randomized controlled trial (RCT) is a planned experiment designed to test the efficacy or effectiveness of an intervention [11]. Use of control conditions is the primary methodology to reduce threats to internal validity in RCTs. The purpose of a control condition is to filter out the variance due to factors that are not specific to the experimental intervention, leaving only the variance due specifically to this treatment [12,13]. A well-designed control condition should maximize our confidence that any difference in the results between the intervention and the control is due to the specific effect of the intervention and not to other general nonspecific factors [14]. The ideal would be to control for: the regression toward the mean, the natural course of the disease, the Hawthorne effect (due to being observed and evaluated, growing the sense of being under medical attention and subsequently growing hope), and the placebo effect (due to belief of being treated, including expectation). In pharmacological studies, the pill placebo is used as a control condition and can control for all above factors. However, the trial design for behavioral interventions has some unique characteristics and control groups vary widely, which influence the effects observed in any given trial. The effects shown for the psychotherapeutic intervention depend on the control condition, and there is indirect evidence suggesting that different control conditions produce different effect sizes [12,13,15]. Depression and anxiety disorders show large placebo responses in comparison with schizophrenia and bipolar disorder [16]. Hope, attention and expectations for improvement play an important role in the treatment of depression [15,17,18]. The control condition could trigger improvement, no response, or even produce a nocebo effect [19]. It is also shown repeatedly that all bona fide psychotherapies, even when used as control conditions, show similar effectiveness [20,21]. On the other hand, many meta-analyses have paid little attention to control conditions in psychotherapeutic trials and may have lumped together various inactive conditions into one comparison [12,22]. However, it is of utmost importance to examine the possible differences among the various inactive control conditions empirically.

Network meta-analysis (NMA) is a recently developed method of evidence synthesis. In network meta-analyses, the information available from within-trial direct comparisons of interventions A and B is combined with indirect comparisons of A and B derived from trials that compare either of the two interventions with a common comparator C (either a third psychological intervention or a control condition), thus enabling more precise estimates of the relative efficacy among alternative interventions (including alternative control conditions) while preserving randomization [23]. NMA permits the comparison of more than two interventions simultaneously (in contrast to pairwise meta-analyses). The synthesis of direct and indirect evidence makes it feasible to compare interventions that would never (or seldom) be compared withy each other in RCTs (such as control conditions). In this way, NMA appears as the best method to accurately estimate the different efficacies of different interventions, as long as transitivity and similarity between these interventions are fulfilled. Establishing which control condition may underestimate or overestimate the effect of psychological interventions through NMA by including several control conditions in the same network simultaneously may touch the very foundation of research into psychological interventions, and may change our interpretation of their empirical evidence base [24,25]. The objective of this study is to examine the influence of the control conditions on effect estimates in psychotherapy trials for depression through NMA. Our hypothesis is that the choice of control condition has an impact on the effect size estimates in RCTs for psychotherapy in depression.

2. Methods

The protocol for this systematic review has been preregistered on the Open Science Foundation platform as part of the overall meta-analysis project (http://www.osf.io/cdfu2). We followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) reporting guideline for NMA [26].

2.1. Criteria for considering studies for this review

2.1.1. Types of participants

2.1.1.1. Inclusion criteria. Participants aged 18 years or older, of both sexes, with a depressive disorder or with elevated levels of depressive symptoms, diagnosed as per operationalized criteria (DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, DSM-5, or ICD-10) or judged so based on a scale used for screening. All studies in which 80% or more of the participants were deemed to satisfy the above criteria were included.

2.1.1.2. Exclusion criteria. We excluded studies recruiting children and adolescents. We also excluded patients who suffer from bipolar depression or psychotic depression.

2.1.2. Types of studies

2.1.2.1. Inclusion criteria. Only RCTs comparing a broadly conceived cognitive behavioral intervention head-

to-head against a control intervention or another cognitive behavioral intervention were included. The psychological intervention had to be implemented at the individual, group, or family level, include face-to-face contact between the patient and the therapist (as opposed to telephone or internet-based interaction between patient and therapist) and consist mainly of verbal communication. Therapies should last at least 4 weeks, but less than 6 months, because this is the threshold for the acute phase. We included trials reported as full-text articles in international peer-reviewed journals and did not include studies reported as abstract only or as dissertation only. No language restrictions and restrictions on publication type were applied.

2.1.2.2. Exclusion criteria. Quasi-RCTs, in which treatment assignment is decided through methods such as alternate days of the week, were excluded. Medical or psychiatric comorbidity was not an exclusion criterion.

Studies in which a pharmacological coadministration was allowed were included, so long as it was administered equally across the arms. The influence of including trials allowing coadministration of drugs was examined in a sensitivity analysis.

2.1.3. Types of interventions

2.1.3.1. Control conditions. Table 1 provides the detailed definitions of the four control conditions of interest for this review.

Care as usual (CAU. Sometimes referred as treatment as usual) is a frequently used control condition in psychotherapy trials. However, the actual contents of CAU differ across settings and across studies [27–29]. Sometimes, no other treatment is available, so patients in the control group receive no treatment at all. Sometimes, a general practitioner gives treatment for depression. Some other times, a qualified psychiatrist is responsible for the treatment and may give antidepressants or/and psychotherapy. In some cases, CAU can be described as ''psychotherapeutic'' or ''nonpsychotherapeutic'' [30]. Such varieties cannot be studied as one node in the network in our present study of control conditions in psychotherapy trials.

We therefore classified what was described as ''treatment as usual'' or ''care as usual'' in the original publication by its contents, where appropriate, as Waiting List (WL,) PsycholPlacebo, or NoTreatment (NT) as per the following rules.

1) If the control condition is used as a waiting list, it is classified as WL.
2) If the trial is CBT + CAU vs. CAU, and this CAU is for depression (e.g., 20% or more patients received antidepressants and/or psychotherapy), we excluded such a trial because it would not contribute to a comparison of CBT against NT or PsycholPlacebo.
3) If the trial is CBT + CAU vs. CAU but this CAU is not for depression (but e.g., for diabetes or hypertension or when this proportion was less than 20% or when it was only information provision), we considered this trial as comparing CBT vs. NT. We conducted a sensitivity analysis to examine the robustness of including such trials.
4) If the trial is CBT vs CAU, we classified this CAU as per its content as NT or PsycholPlacebo.

2.1.3.2. Active interventions. Studies with an active drug as active intervention were not eligible. Clinical heterogeneity within therapeutic approaches is an issue that often arises in psychotherapeutic studies [10]. To reduce this problem, we focused on broadly conceived cognitive or behavioral psychotherapies, as proposed by an expert taxonomy of psychotherapy for depression [10,14,31], including cognitive behavioral therapy (CBT), behavioral activation (BA), problem-solving therapy, and 3rd wave CBT (3W). Table 1 provides the definitions for respective interventions.

2.1.4. Types of outcomes

The primary outcome was the response, defined as 50% or greater reduction in depressive symptomatology from baseline to the end of the study. The secondary outcome was the depression severity at end of treatment. When only change scores were provided, we pooled them with end point scores [32]. The outcome was assessed at the end of the acute phase treatment for every study. If more than one assessment times were reported, we extracted the last assessment immediately after the acute phase intervention (between 4 weeks and 6 months). Follow-up was not included. For each comparison, the effect size indicating the difference between the two arms was expressed as the odds ratio (OR) [33] or as standardized mean difference corrected for small sample size [34]. We included only validated depression measures. However, when more than one measure was used in a study, we prioritized the scales in the following order to minimize potential selective outcome reporting bias: 1) Hamilton Rating Scale for Depression (any version), 2) Beck Depression Inventory I or II, 3) another clinician-rated instrument, or 4) another self-report instrument.

2.2. Data collection

2.2.1. Search methods for identification of studies

We based the present study on the database of randomized trials of psychotherapies for depression [35]. This database has been continuously updated by a comprehensive literature search up to January 1, 2019 of four major bibliographical databases (PubMed, PsycINFO, Embase, and the Cochrane Library). The search used a combination of index and text terms indicative of depression and psychotherapies, with filters for RCTs. Appendix A gives the full search string for one database (PubMed). Furthermore, we checked the references of earlier meta-analyses on psychological treatments of depression.

Table 1. Definitions of control conditions and cognitive behavioral interventions included in this review

Controls and interventions	Description/definition
No psychological treatment (NoTreatment)	Participants receive assessment only, with or without simple provision of informational material and/or minimal therapist contact, and they know that they will not receive the active treatment in question after the trial.
Waiting list (WL)	Participants receive assessment, with or without simple provision of informational material and/or minimal therapist contact, and they know that they will receive the active treatment in question after the waiting phase.
Psychological placebo (PsychoIPla)	Participants receive a face-to-face inactive intervention that can be perceived effective by the participants. They spend time with the therapist of the same duration and frequency with the experimental treatment, but no specific therapeutic techniques are administered. Supportive counseling will be included under this category if it is intended to serve as a control condition as defined previously in the design of the study.
Pill placebo (PillPla)	Participants receive an inactive pill. This usually happens when a multiarm trial is conducted.
Cognitive behavioral therapy (CBT)	In CBT, the therapists focus on the impact that a patient's present dysfunctional thoughts have on current behavior and future functioning. CBT is aimed at evaluating, challenging, and modifying a patient's dysfunctional beliefs (cognitive restructuring). In this form of treatment, the therapist mostly emphasizes homework assignments and outside-of-session activities. Therapists exert an active influence over therapeutic interactions and topics of discussion, use a psycho educational approach, and teach patients new ways of coping with stressful situations. The most used subtypes are CBT according to Beck's manual (Beck et al., 1979) and the ''Coping with Depression'' course (Lewinsohn et al., 1984).
Behavioral activation therapy (BA)	We considered an intervention to be behavioral activation when the registration of pleasant activities and the increase of positive interactions between a person and his or her environment were the core elements of the treatment. Social skills training could be a part of the intervention. There are several subtypes of behavioral activation (Mazzucchelli et al. 2009).
Problem solving therapy (PST)	We defined PST as a psychological intervention in which the following elements had to be included: definition of personal problems, generation of multiple solutions to each problem, selection of the best solution, the working out of a systematic plan for this solution, and evaluation as to whether the solution has resolved the problem. Subtypes of PST are described elsewhere (Cuijpers et al., 2018).
Third wave cognitive behavioral therapies (3W)	Third wave therapies are a heterogeneous group of therapies that introduce several new techniques to cognitive behavior therapies. They have in common that they abandon or only cautiously use content-oriented cognitive interventions and the use of skills deficit models to delineate the core maintaining mechanisms of the addressed disorders (Kahl, Winter, & Schweiger, 2012). Well-known therapies that we clustered in this category include acceptance and commitment therapy, mindfulness-based CBT, and metacognitive therapy.

2.2.2. *Selection of studies*

Two independent review authors (EK, PC, MC, and CM) examined the abstracts of all publications obtained through the search strategy. Full articles of all studies identified by at least one of the two review authors were then obtained and inspected by the same review authors. Conflicts of opinion regarding eligibility of a study were discussed, having retrieved the full article and consulted the authors if necessary, until consensus was reached. Methodological experts (TAF and PC) were consulted if necessary.

2.2.3. *Data extraction and management*

Data from each study were extracted by two independent review authors (IM, SK, AO, and EGO). Any disagreement was resolved through discussion and in consultation with the principal investigators. Where necessary, the authors of the studies were contacted for further information.

2.2.4. *Assessment of risk of bias in included studies*

Two independent review authors (EK, PC, MC, and CM) assessed the risk of bias in selected studies. Any disagreement was resolved through discussion and in consultation with the principal investigators (PC and TAF).

Risk of bias was assessed for each included study using the Cochrane Collaboration risk of bias tool as a model [36]. The following domains were considered: adequate generation of random sequence, allocation concealment, blinding of outcome assessors, and dealing with incomplete outcome data. A study was rated as low overall risk of bias when all four items were rated as low risk, as moderate

overall risk of bias when three or two items were rated at low risk, and as high overall risk of bias when only one or no item was rated at low risk.

2.2.5. Dealing with missing data

When the response rate was not reported, we imputed it using the mean at baseline and the mean and SD at end point [37]. Patients randomized but not included in the original analyses were assumed to be nonresponders and included in the current analyses following the intention-to-treat principle. Missing continuous outcome data were analyzed on an end point basis, by using the last observation carried forward (LOCF) to the final assessment if LOCF data were reported by the trial authors or by including only participants with a final assessment. When *P*-values, t-values, confidence intervals, or standard errors were reported in articles, SD was calculated from their values.

2.3. Data synthesis

We conducted NMA to synthesize the available evidence from the entire network of trials by integrating direct and indirect estimates for each comparison into a single summary treatment effect. We used the package *netmeta* (version 1.2-0) of R (version 3.6.1) for all analyses, except for *gemtc* (version 0.8-2) and JAGS (version 4.3.0) for metaregression analyses to investigate sources of heterogeneity and OpenBUGS (version 3.2.3) for metaregression analyses adjusting for small study effects between interventions and control conditions [38].

2.3.1. Assessment of statistical heterogeneity

We assessed statistically the presence of heterogeneity using the I-squared statistic and its 95% confidence interval that measures the percentage of variability that cannot be attributed to random error. The degree of statistical heterogeneity in the entire network was assessed through the magnitude of the heterogeneity variance parameter (τ^2) estimated from the NMA models. τ^2 was compared with the empirical distribution as derived for dichotomous and continuous outcomes [39,40].

2.3.2. Assessment of transitivity and statistical inconsistency

Transitivity is the basic prerequirement of NMA: it requires that potentially important effect modifiers are evenly distributed among the comparisons. We assessed transitivity by examining the average age, proportion of women, recruitment settings, use of operational diagnostic criteria, and number of sessions among comparisons which had five or more included studies.

Consistency in the network (i.e., direct and indirect evidence are in agreement) is statistical expression of transitivity: it can be violated either in the entire network or in certain parts (i.e., loops of evidence) of the network. We evaluated the presence of inconsistency using the following approaches:

2.3.2.1. Global tests. To check the assumption of consistency in the entire network simultaneously we used the ''design-by-treatment'' model [41]. This method accounts for different sources of inconsistency that can occur when studies with different designs (two-arm trials vs. three-arm trials) give different results as well as disagreement between direct and indirect evidence. Using this approach, we inferred about the presence of inconsistency from any source in the entire network based on a chi-square test.

2.3.2.2. Local tests. To evaluate the presence of inconsistency locally, we used the side-splitting approach [42]. This method evaluates the consistency assumption by statistically comparing the direct estimate and the indirect estimate for each mixed estimate.

2.3.3. Investigation of sources of heterogeneity and inconsistency

We explored possible sources of heterogeneity and inconsistency [43] through metaregression using the potential effect modifiers listed below.

1. Year of publication
2. Blinding of outcome assessors

2.3.4. Investigation of small study effects

Small study effects including publication bias may be expected between active treatments and controls. We examined this by drawing funnel plots between all control conditions and all active interventions.

2.3.5. Sensitivity analyses

We conducted the following three sensitivity analyses.

1. Treating all active psychotherapies as one node in the network, given little evidence for differences in efficacy among active psychotherapies [10,14].
2. Excluding studies which examined psychotherapies among patients with physical comorbidities because in such trials face-to-face interventions targeting physical comorbidities were administered to both active arms and control arms and may have dampened the possible differences between them.
3. Excluding studies at high risk of bias.
4. If small study effects were observed, adjusting for study variance for comparisons with suspected small study effects.

2.4. Changes from the protocol

The primary outcome was changed from the continuous outcome of depression severity to the dichotomous outcome of response in depression. The dichotomous

outcome was preferred because it enables closer adherence to the intention-to-treat principle by conservatively assuming all dropouts to be no-responders and also because the obtained results are more clinically interpretable [44]. The analyses for the continuous outcome, however, were reported as the secondary outcome.

In the primary network, we differentiated different forms of cognitive and/or behavioral therapies. This was performed to increase the connectedness of the network, so that the mixed estimates would be more stable and more consistent. However, a NMA lumping all broadly conceived CBT interventions into one node was reported as a sensitivity analysis.

The prespecified sensitivity analysis to exclude studies which had allowed coadministration of drugs could not be conducted as most of the study reports were not clear if they had allowed such and if the patients in fact received such during the original randomized trials. We added one sensitivity analysis excluding studies at high risk of bias, as requested by one of the reviewers.

3. Results

3.1. Included studies

After examining 22,049 abstracts (16,701 after removal of duplicates), we assessed 2,553 full-text articles for eligibility, in which we identified 308 RCTs of psychotherapies comparing one against another or against a control condition. By limiting the psychotherapies to broadly defined cognitive behavioral ones and by scrutinizing the contents of the control conditions, we finally arrived at 123 studies (12,596 participants) for inclusion in the current NMA. Appendix B depicts the PRISMA flowchart, Appendix C presents the references, and Appendix D presents the selected characteristics of the included studies. Appendix D also provides detailed descriptions of interventions judged to constitute PsycholPlacebo and NoTreatment, with citations from the original publications to support our judgments.

3.2. Characteristics of the included studies and participants

Table 2 presents the aggregated characteristics of the included studies and the participants. The 123 included studies were published between 1977 and 2018 (median: 2009). The participants were in their 40s on average, and 69% were women. Approximately, half were diagnosed with major or minor depression, whereas the others were selected based on a cutoff score of a rating scale.

Appendix D provides the risk of bias assessments for individual studies. Overall, 31 studies (25.2%) were considered to be at low overall risk of bias, 49 (39.8%) at moderate risk, and 43 (35.0%) at high risk of bias.

3.3. Network meta-analyses

Figure 1 presents the network diagram of the included studies for the primary outcome. Among the control conditions of interest, there was only one direct comparison between WL and PsycholPlacebo. However, overall, the network was well connected, without any singly standing node, and every control condition was compared at least with three active intervention nodes.

Table 3 shows the league table among all psychotherapies and control conditions for response (lower left triangle) and depression severity (upper right triangle). Figure 2 shows the ranked forest plot, with WL as a reference. There were no statistically significant differences among the active interventions, which were all significantly superior to all control conditions (except for 3W over PillPlacebo). Among the control conditions, PillPlacebo, NoTreatment, and PsycholPlacebo were all significantly superior to WL. The effect sizes of psychotherapies, therefore, were substantively different depending on which control condition was used.

3.4. Investigation of heterogeneity and inconsistency

I-squared of the network was 55.9% (95% CI: 45.9%; to 64.0%) and suggested moderate heterogeneity. Tau-squared, the common heterogeneity parameter, was 0.35, well within the 95% prediction interval for tau-squared (0.005−3.33) for subjective outcomes in nonpharmacological interventions [45].

Potential effect modifiers were apparently evenly distributed across comparisons, especially when there were more than 10 studies included per comparison, suggesting that the transitivity of the network was maintained. The design-by-treatment global test of inconsistency was not significant ($P = 0.44$). None of the side-splitting tests suggested inconsistency between direct and indirect estimates beyond chance (Appendix E).

The funnel plot of comparisons between active psychotherapies and various control conditions was highly asymmetric (Appendix F, Egger's $P < 0.001$).

None of the meta-regressions by year of the study or blinding of assessor was statistically significant, with estimated coefficients of −0.42 (−1.22 to 0.37) and 0.86 (−0.01 to 1.75), respectively.

3.5. Sensitivity analyses

We first attempted a prespecified sensitivity analysis lumping all broadly conceived CBT arms into one node. However, the network (Appendix G) turned out to be disconnected, and we deemed it not worthwhile to run NMA on this network.

Appendix H shows the results of the other three sensitivity analyses. When we excluded studies targeting patients with comorbid physical diseases or studies with

Table 2. Aggregate characteristics of the included studies

	N	%	Mean	SD or range
Participants				
Mean age			45.4	(Range: 18.2-81.9)
Proportion of women			0.69	(Range: 0-1)
Recruitment				
Community	63	51.2%		
Clinical	23	18.7%		
Other	37	30.1%		
Target group				
Adults in general	46	37.4%		
Older adults	15	12.2%		
Students	12	9.8%		
Perinatal depression	7	5.7%		
General medical disorder	30	24.4%		
Other	13	10.6%		
Diagnosis				
Major depression	38	30.9%		
Major and minor depression	27	21.9%		
Cutoff score	51	41.5%		
Subclinical depression	6	4.9%		
Chronic depression	1	0.8%		
Interventions				
Number of sessions			10.0	(Range: 4-24)
Intervention format				
Individual	110	42.3%[a]		
Group	108	41.5%[a]		
Other	42	16.2%[a]		
Active intervention arms				
CBT	94	36.1%[a]		
BA	19	7.3%[a]		
PST	16	6.2%[a]		
3W	17	6.5%[a]		
Control arms				
WL	57	21.9%[a]		
NoTreatment	32	12.3%[a]		
PsycholPlacebo	17	6.5%[a]		
PillPlacebo	8	3.1%[a]		
Outcomes				
Instruments used				
HAMD	43	35.0%		
BDI	29	23.6%		
BDI-II	16	13.0%		
CES-D	11	8.9%		
PHQ-9	6	4.9%		
Others	18	14.6%		
Study				
Year of the study			2006	(Median 2009; Range: 1977-2018)
Number of participants per arm			48.4	55.6 (Range: 7-463)
Country				
USA	50	40.6%		
UK	12	9.8%		

(*Continued*)

Table 2. Continued

	N	%	Mean	SD or range
EU	23	18.7%		
Canada	6	4.9%		
Australia	9	7.3%		
East Asia	9	7.3%		
Other	14	11.4%		
Risk of bias (low)				
Sequence generation	56	45.5%		
Allocation concealment	50	40.6%		
Blinding of assessor	41	33.3%		
ITT analysis	65	52.8%		

[a] The denominator is the total number of arms (not the total number of studies).

high risk of bias, the results were essentially similar to the primary results. When we adjusted for small study effects between active interventions and control conditions through metaregression, the effect sizes became much smaller and only BA, CBT, and 3W were significantly more effective than WL: however, the ranking of the interventions and the control conditions remained the same.

4. Discussion

4.1. Principal findings

In this NMA, we assessed comparative data from 123 randomized trials of psychotherapies in the acute phase of depression. The network was well connected, homogenous, and consistent. Broadly conceived CBT therapies were used as active interventions. We found that there are important differences among the various control conditions typically used in psychotherapy trials. Different control conditions lead to different estimates of efficacy for the same intervention. Among them, WL appears to be the weakest, that is, an active intervention will have the biggest effect size when the control condition is WL. By contrast, PillPlacebo appears to be the strongest control condition, with PsycholPlacebo and NoTreatment coming inbetween.

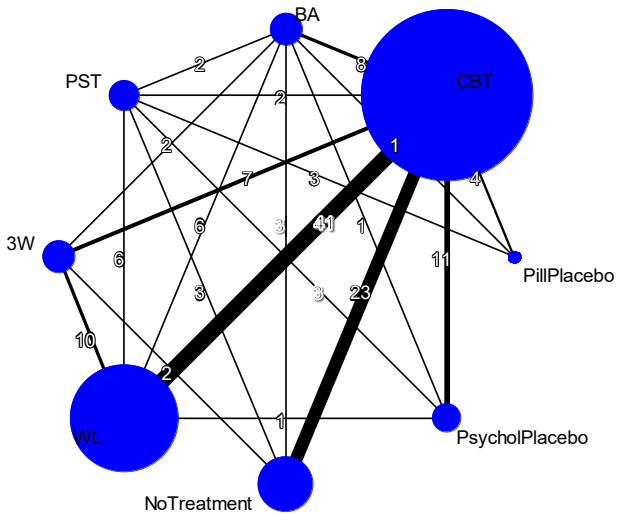

Fig. 1. Network diagram for response.

4.2. Strengths and limitations of the study

To our knowledge, it is the first NMA comparing all typical control conditions including WL, NoTreatment, PsycholPlacebo, and PillPlacebo simultaneously in the NMA. The identified studies fulfilled two basic assumptions for NMA: it had moderate heterogeneity (I^2 = 55.9%) and there was no evidence for inconsistency in the network. Using the rigorous definitions for control conditions, we were still able to include 123 studies (12,596 participants) in the network, enabling more precise estimates for relative efficacies among various control conditions.

Furthermore, we have included studies that are not limited to western countries, but from all continents, from different health systems and different cultures; this makes our findings more comprehensive. We did not include the term "CAU" ("Care as usual" or "Treatment as usual") as its contents vary enormously, and it produces great heterogeneity [27,29]. CAU can range from an intervention with antidepressants plus psychotherapy to no intervention at all, depending on the settings and available resources [46]. In an effort to reduce the heterogeneity of CAU, Wampold et al. (2011) suggested the use of a scale to classify CAU as "psychotherapeutic" or not [30]. We classified CAU as per the actual treatments given to patients, which reduced heterogeneity and allowed us to focus on comparisons among rigorously defined control conditions.

Compared with psychopharmacological trials, psychotherapeutic trials appear with higher risk of bias because there is no way for a psychotherapeutic trial to be double

Table 3. League table for response (lower left) and for depression severity (upper right)

BA	−0.03 (−0.51 to 0.45)	−0.03 (−0.37 to 0.30)	−0.12 (−0.55 to 0.32)	−0.66 (−1.31 to 0.00)	−1.04 (−1.52 to −0.56)	−0.68 (−1.07 to −0.29)	−1.27 (−1.64 to −0.91)
1.00 (0.57 to 1.78)	PST	−0.01 (−0.40 to 0.39)	−0.09 (−0.58 to 0.40)	−0.63 (−1.29 to 0.04)	**−1.01 (-1.47 to -0.55)**	**−0.65 (-1.07 to -0.22)**	**−1.24 (-1.65 to -0.83)**
1.15 (0.75 to 1.77)	1.15 (0.72 to 1.82)	CBT	−0.08 (−0.40 to 0.24)	**−0.62 (-1.22 to -0.02)**	**−1.00 (-1.37 to -0.64)**	**−0.64 (-0.89 to -0.39)**	**−1.24 (-1.43 to -1.04)**
1.25 (0.71 to 2.22)	1.25 (0.67 to 2.31)	1.09 (0.71 to 1.68)	3W	−0.54 (−1.21 to 0.14)	**−0.92 (-1.40 to -0.44)**	**−0.56 (-0.95 to -0.17)**	**−1.15 (-1.48 to -0.83)**
2.03 (1.05 to 3.93)	**2.02 (1.13 to 3.61)**	**1.77 (1.01 to 3.09)**	1.62 (0.81 to 3.26)	PillPla	−0.38 (−1.07 to 0.31)	−0.02 (−0.66 to 0.62)	−0.61 (−1.24 to 0.01)
2.67 (1.46 to 4.89)	**2.66 (1.52 to 4.66)**	**2.32 (1.48 to 3.63)**	**2.13 (1.15 to 3.95)**	1.31 (0.66 to 2.60)	PsycholPla	0.36 (−0.07 to 0.79)	**−0.23 (-0.64 to 0.17)**
2.81 (1.75 to 4.52)	**2.80 (1.69 to 4.64)**	**2.44 (1.81 to 3.30)**	**2.24 (1.35 to 3.74)**	1.38 (0.75 to 2.55)	1.05 (0.62 to 1.78)	NoTreatment	**−0.59 (-0.90 to -0.29)**
5.42 (3.35 to 8.77)	**5.39 (3.25 to 8.95)**	**4.71 (3.59 to 6.17)**	**4.32 (2.77 to 6.75)**	**2.66 (1.45 to 4.89)**	**2.03 (1.21 to 3.39)**	**1.93 (1.30 to 2.86)**	WL

In the lower left triangle, OR > 1 means that the column-defining intervention increases the response in comparison with the row-defining intervention. In the upper right triangle, SMD < 0 means that the row-defining intervention decreases the depression severity more than the column-defining intervention.

Bold indicates statistically significant effect sizes.

blind. Nevertheless, in this NMA, only RCTs were included, lowering the risk of bias; about two-thirds of the studies had low to moderate risk of bias. We included only face to face interventions: in this way, the placebo effect was homogeneous for all the comparisons examined. We excluded the studies that had more than 20% coadministration with drugs, which diminished confounding by the presence of antidepressants.

Another important issue is publication bias, whereby the "negative" results are less likely to be published. Unfortunately, there is no established method to adjust for such small study effects. We found significant asymmetry in the funnel plot of comparisons between active psychotherapies vs. control conditions (Appendix F). If the small study effects existed to the same degree only between active interventions, on the one hand, and control conditions, on the other, the comparative effectiveness among the first group and among the second group, respectively, would not be affected. A sensitivity analysis metaregression for variance in the comparisons between active interventions and controls expectedly reduced the effect sizes of the former over the latter and also among the former or the latter arms themselves. However, the relative rankings remained the same.

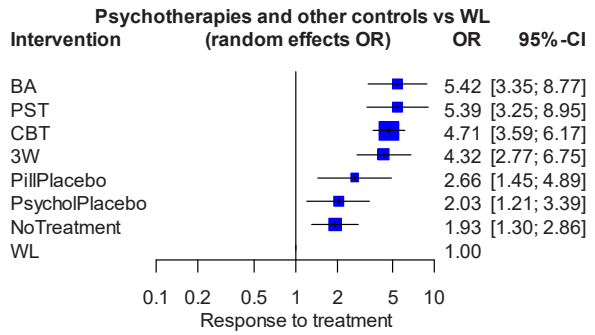

Fig. 2. Ranked forest plots for response in the primary analysis.

Last but not least, this NMA is not exempt from the limitations of the original studies in it: for example, the participants were those who had consented to randomized research and may not be representative of those who would seek psychotherapies; the interventions and their therapists may be better controlled and of higher quality than those actually administered in the real world; the therapists and their supervisors were not randomized within trials and there may be confounding between interventions and therapists.

4.3. Comparison with other studies

The role of control conditions in psychotherapeutic trials has been investigated before. Mohr et al. found differences among control conditions and suggested that the choice of the control condition has an impact on the effect sizes found between the active treatment and the control condition [12]. However, great heterogeneity was present, and they did not use a NMA, but they performed multiple meta-analyses, making the synthesis of the results uncertain. They had included CAU as a control condition, without distinguishing its elements and found that CAU produced similar effect sizes to WL and NT. Considering WL vs. NT, they found that WL presented greater effect sizes for active psychotherapies than NT, something that is in line with our findings.

Furukawa et al. applied a NMA including CBT as the active intervention and WL, NT, and PsycholPlacebo as

control conditions [19]. They had also excluded CAU, to reduce heterogeneity, but they had not included pill placebo, which appears in our analysis as a strong control condition. They reported WL as a ''nocebo condition'' because it produced bigger effect sizes than PsycholPlacebo and even NT. The present study has more than doubled the sample size and has been able to lead to firmer inferences about relative efficacies among various control conditions.

Barth et al. performed a NMA comparing the efficacy of psychotherapeutic interventions for depression [14]. Apart from their main findings, they also concluded that different control conditions produce different effect sizes. They used CAU as a control condition, although they reported that CAU is a factor that can confound the results. They also lumped psychological placebo and pill placebo as one control condition, whereas in our study, the latter appeared stronger than the former.

Munder et al. [21] were based on data from Cuijpers et al. [47] and tried to distinguish the efficacy of different psychotherapeutic approaches for depression. Following Cuijpers et al. methodology, they included CAU and ''other'' controls in their analyses. They also performed multiple separate analyses and excluded outliers, ''based on visual inspection.'' Apart from their main outcomes, they examined control conditions as well. But they did not test the differences of control conditions among each other. They also considered WL as the same with NT, whereas in our study, these two conditions differ (in theory and in ''effectiveness''). They concluded that WL is not a ''nocebo'' control condition but it could reflect the natural course of depression.

Depression is a disorder with a large placebo response [16,17]. Being under medical attention, even if this is not followed by medical interventions, patients have more hope and expectations (allegiance or Hawthorne effect). Hope for improvement is a crucial part for treatment response. The choice of the control condition can affect the outcome of an RCT because some control conditions, such as pill placebo, can produce effects comparable with active treatments [48,49], while some others, such as WL, can produce neutral or even negative effects, leading to larger differences between interventions and controls [15,19]. WL seems to be ''less effective'' than NT. This may be explained from the design of studies: patients who are randomized in NT or WL are free to seek other treatments while they are monitored. However, patients on WL are more likely to wait for the active treatment, whereas patients on NT have nothing to lose and may be more active to seek for help elsewhere. WL may also increase a state of helplessness during the waiting period, while increasing expectations for the upcoming intervention. It has been proposed that different control conditions could be used in different stages of RCTs [13,15]. The ''weakest,'' such as WL, could be used at the initial phase when a new intervention is introduced, and the ''strongest,'' such as pill placebo, could be used at the final phase when the intervention is considered for inclusion in treatment guidelines.

5. Conclusions and policy implications

In summary, we found that different control conditions produce different effect sizes in psychotherapy RCTs for depression when they are all considered simultaneously in NMA. All control conditions (PillPlacebo, NT, and PsycholPlacebo) had higher ORs for response against WL. The findings have clear implications first for the interpretations of RCTs of psychotherapies, second for the design of such RCTs, and finally for the conduct of their meta-analyses. First, the effect sizes obtained in RCTs must be interpreted differently depending on the control condition used. Second, one may use weaker controls only in the earlier phases of therapy evaluations but must move to stronger controls in confirmatory studies. And finally, in meta-analyses and network meta-analyses of psychotherapy trials, one should no longer lump various control conditions into one comparison group.

Acknowledgments

This research was supported in part by Matsumae Foundation Research Fellowship 2019 to IM, and JSPS, Japan Grant-in-Aid for Scientific Research (Grant Number 17K19808) to TAF. EO is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2020.12.012.

References

[1] World Health Organization. Depression and other common mental disorders: Global Health Estimates 2017. Geneva: World Health Organization; 2017: CC BY-NC-SA 3.0 IGO.
[2] GBD 2015 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 315 diseases and injuries and healthy life expectancy (HALE), 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016;388:1603−58.
[3] Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013;10:e1001547.
[4] Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the Community from 30 countries between 1994 and 2014. Sci Rep 2018;8:2861.
[5] Lucht M, Schaub RT, Meyer C, Hapke U, Rumpf HJ, Bartels T, et al. Gender differences in unipolar depression: a general population survey of adults between age 18 to 64 of German nationality. J Affect Disord 2003;77:203−11.

[6] Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2197−223.
[7] Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. World Psychiatry 2020;19:92−107.
[8] Chambless DL, Hollon SD. Defining empirically supported therapies. J Consult Clin Psychol 1998;66:7−18.
[9] Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. Int Rev Psychiatry 2011;23:318−27.
[10] Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. Psychother Res 2020;30:279−93.
[11] Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. Health Psychol 2004;23:443−51.
[12] Mohr DC, Ho J, Hart TL, Baron KG, Berendsen M, Beckner V, et al. Control condition design and implementation features in controlled trials: a meta-analysis of trials evaluating psychotherapy for depression. Transl Behav Med 2014;4:407−23.
[13] Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, et al. The selection and design of control conditions for randomized controlled trials of psychological interventions. Psychother Psychosom 2009;78:275−84.
[14] Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. Plos Med 2013;10:e1001454.
[15] Gold SM, Enck P, Hasselmann H, Friede T, Hegerl U, Mohr DC, et al. Control conditions for randomised trials of behavioural interventions in psychiatry: a decision framework. Lancet Psychiatry 2017;4:725−32.
[16] Khan A, Kolts RL, Rapaport MH, Krishnan KRR, Brodhead AE, Browns WA. Magnitude of placebo response and drug-placebo differences across psychiatric disorders. Psychol Med 2005;35:743−9.
[17] Furukawa TA, Cipriani A, Atkinson LZ, Leucht S, Ogawa Y, Takeshima N, et al. Placebo response rates in antidepressant trials: a systematic review of published and unpublished double-blind randomised controlled studies. Lancet Psychiatry 2016;3:1059−66.
[18] Hegerl U, Allgaier A-K, Henkel V, Mergl R. Can effects of antidepressants in patients with mild depression be considered as clinically significant? J Affect Disord 2012;138:183−91.
[19] Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, et al. Waiting list may be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. Acta Psychiatr Scand 2014;130:181−92.
[20] Wampold BE, Mondin GW, Moody M, Stich F, Benson K, Ahn H. A meta-analysis of outcome studies comparing bona fide psychotherapies: Empiricially, ‘‘all must have prizes. Psychol Bull 1997;122:203−15.
[21] Munder T, Flückiger C, Leichsenring F, Abbass AA, Hilsenroth MJ, Luyten P, et al. Is psychotherapy effective? A re-analysis of treatments for depression. Epidemiol Psychiatr Sci 2019;28:268−74.
[22] Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. Psychol Med 2010;40:211−23.
[23] Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. Ann Intern Med 2013;159:130−7.
[24] Cuijpers P, Turner EH, Mohr DC, Hofmann SG, Andersson G, Berking M, et al. Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis. Psychol Med 2014;44:685−95.
[25] Furukawa TA, Schramm E, Weitz ES, Salanti G, Efthimiou O, Michalak J, et al. Cognitive-Behavioural Analysis System of Psychotherapy (CBASP), a drug, or their combination: differential therapeutics for persistent depressive disorder: a study protocol of an individual participant data network meta-analysis. BMJ Open 2016;6:e011769.
[26] Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. Ann Intern Med 2015;162:777−84.
[27] Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. J Affect Disord 2015;175:152−67.
[28] Witt K, de Moraes DP, Salisbury TT, Arensman E, Gunnell D, Hazell P, et al. Treatment as usual (TAU) as a control condition in trials of cognitive behavioural-based psychotherapy for self-harm: impact of content and quality on outcomes in a systematic review. J Affect Disord 2018;235:434−47.
[29] Cuijpers P, Quero S, Papola D, Cristea IA, Karyotaki E. Care-as-usual control groups across different settings in randomized trials on psychotherapy for adult depression: a meta-analysis. Psychol Med 20191−11.
[30] Wampold BE, Budge SL, Laska KM, Del Re AC, Baardseth TP, Fluckiger C, et al. Evidence-based treatments for depression and anxiety versus treatment-as-usual: a meta-analysis of direct comparisons. Clin Psychol Rev 2011;31:1304−12.
[31] Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. J Consult Clin Psychol 2008;76:909−22.
[32] da Costa BR, Nüesch E, Rutjes AW, Johnston BC, Reichenbach S, Trelle S, et al. Combining follow-up and change data is valid in meta-analyses of continuous outcomes: a meta-epidemiological study. J Clin Epidemiol 2013;66:847−55.
[33] Bakbergenuly I, Hoaglin DC, Kulinskaya E. Pitfalls of using the risk ratio in meta-analysis. Res Synth Methods 2019;10:398−419.
[34] Hedges L, Olkin I. Statistical methods in meta-analysis. Stat Med 1985;20. https://doi.org/10.2307/1164953.
[35] Cuijpers P, Karyotaki E, Ciharova M. A meta-analytic database of randomised trials on psychotherapies for depression 2019. https://doi.org/10.17605/OSF.IO/825C6.
[36] Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions. vol. 5.1.0. 2011. Available at www.cochrane-handbook.org.
[37] Furukawa TA, Cipriani A, Barbui C, Brambilla P, Watanabe N. Imputing response rates from means and standard deviations in meta-analyses. Int Clin Psychopharmacol 2005;20:49−52.
[38] Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Methods 2012;3:161−76.
[39] Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. J Clin Epidemiol 2015;68:52−60.
[40] Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41:818−27.
[41] Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods 2012;3:98−110.
[42] Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010;29:932−44.
[43] Salanti G, Marinho V, Higgins JPT. A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol 2009;62:857−64.

[44] Furukawa TA, Salanti G, Atkinson LZ, Leucht S, Ruhe HG, Turner EH, et al. Comparative efficacy and acceptability of first-generation and second-generation antidepressants in the acute treatment of major depression: protocol for a network meta-analysis. BMJ Open 2016;6:e010919.

[45] Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41:818−27.

[46] Freedland KE, Mohr DC, Davidson KW, Schwartz JE. Usual and unusual care: existing practice control groups in randomized controlled trials of behavioral interventions. Psychosom Med 2011;73:323−35.

[47] Cuijpers P, Karyotaki E, Reijnders M, Ebert DD. Was Eysenck right after all? A reassessment of the effects of psychotherapy for adult depression. Epidemiol Psychiatr Sci 2019;28:21−30.

[48] Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. Br J Psychiatry 2017;210:190−6.

[49] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252−60.